

```

1  TGCTGGGGCA CCTGAAGGAG ACTTGGGGGC ACCCGCGTCG TGCCTCCTGG
51 GTTGTGAGGA GTCGCCGCTG CCGCCACTGC CTGTGCTTCA TGAGGAAGAT
101 GCTCGCCGCC GTCTCCCGCG TGCTGTCTGG CGCTTCTCAG AAGCCGGCAA
151 GCAGAGTGCT GGTAGCATCC CGTAATTTTG CAAATGATGC TACATTTGAA
201 ATTAAGAAAT GTGACCTTCA CCGGCTGGAA GAAGGCCCTC CTGTCACAAC
251 AGTGCTCACC AGGGAGGATG GGCTCAAATA CTACAGGATG ATGCAGACTG
301 TACGCCGAAT GGAGTTGAAA GCAGATCAGC TGTATAAACA GAAAATTATT
351 CGTGGTTTCT GTCACCTGTG TGATGGTCAG TTTCTCCTTC CTCTAACACA
401 GGAAGCTTGC TGTGTGGGCC TGGAGGCCGG CATCAACCCC ACAGACCATC
451 TCATCACAGC CTACCGGGCT CACGGCTTTA CTTTCACCCG GGGCCTTTCC
501 GTCCGAGAAA TTCTCGCAGA GCTTACAGGA CGAAAAGGAG GTTGTGCTAA
551 AGCGAAAAGGA GGATCGATGC ACATGTATGC CAAGAACTTC TACGGGGGCA
601 ATGGCATCGT GGGAGCGCAG GTGCCCCCTG GCGCTGGGAT TGCTCTAGCC
651 TGTAAGTATA ATGGAAGAAG TGAGGTCTGC CTGACTTTAT ATGGCGATGG
701 TGCTGCTAAC CAGGGCCAGA TATTCGAAGC TTACAACATG GCAGCTTTGT
751 GGAAATTACC TTGTATTTTC ATCTGTGAGA ATAATCGCTA TGAATGGGA
801 ACGTCTGTTG AGAGAGCGGC AGCCAGCACT GATTACTACA AGAGAGGCGA
851 TTTCAATTCT GGGCTGAGAG TGGATGGAAT GGATATCCTG TGCCTCCGAG
901 AGGCAACAAG GTTTGCTGCT GCCTATTGTA GATCTGGGAA GGGGCCCATC
951 CTGATGGAGC TGCAGACTTA CCGTTACCAC GGACACAGTA TGAGTGACCC
1001 TGGAGTCAGT TACCGTACAC GAGAAGAAAT TCAGGAAGTA AGAAGTAAGA
1051 GTGACCCAT TATGCTTCTC AAGGACAGGA TGGTGAACAG CAATCTTGCC
1101 AGTGTGGAAG AACTAAAGGA AATTGTATGT GAAGTGAGGA AGGAGATTGA
1151 GGATGCTGCC CAGTTTGCCA CGGCCGATCC TGAGCCACCT TTGGAAGAGC
1201 TGGGCTACCA CATCTACTCC AGCGACCCAC CTTTGAAGT TCGTGGTGCC
1251 AATCAGTGGA TCAAGTTTAA GTCAGTCAGT TAAGGGGAGG AGAAGGAGAG
1301 GTTATACCTT CAGGGGGCTA CCAGACAGTG TTCTCAACTT GGTTAAGGAG
1351 GAAGAAAACC CAGTCAATGA AATTCAATGA AATTCTTGGA AACTTCATT
1401 AAGTGTGTAG ATTGAGCAGG TAGTAATTGC ATGCAGTTT TACATTAGTG
1451 CATTAAAGA TGAATTATTG AGTGCTTAAA AAAAAAAAAA AAAAAAAAAA
1501 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA (SEQ ID NO:1)

```

#### FEATURES:

5'UTR: 1-89  
Start Codon: 90  
Stop Codon: 1281  
3'UTR: 1284

#### Homologous proteins:

##### Top 10 BLAST Hits

			Score	E
CRA	18000004925454 /altid=gi 387011 /def=gb AAA60055.1  (J03503...		846	0.0
CRA	18000004920128 /altid=gi 4505685 /def=ref NP_000275.1  pyru...		793	0.0
CRA	18000004938217 /altid=gi 6679261 /def=ref NP_032836.1  pyru...		783	0.0
CRA	18000004939896 /altid=gi 66035 /def=pir  DERTP1 pyruvate de...		782	0.0
CRA	18000004949905 /altid=gi 129064 /def=sp P26284 ODPA_RAT PYR...		779	0.0
CRA	18000004885327 /altid=gi 266686 /def=sp P29804 ODPA_PIG PYR...		777	0.0
CRA	18000004969398 /altid=gi 448580 /def=prf  1917268A pyruvate...		729	0.0
CRA	18000005012775 /altid=gi 1079460 /def=pir  A49360 pyruvate ...		718	0.0
CRA	18000004884262 /altid=gi 1709452 /def=sp P52900 ODPA_SMIMA ...		709	0.0
CRA	18000004925713 /altid=gi 4885543 /def=ref NP_005381.1  pyru...		680	0.0

FIGURE 1A

BLAST hits to dbEST :

gi 10991237	/dataset=dbest	/taxon=96...	1354	0.0
gi 14051054	/dataset=dbest	/taxon=960...	1415	0.0
gi 14076211	/dataset=dbest	/taxon=960...	1382	0.0
gi 11251518	/dataset=dbest	/taxon=96...	1340	0.0
gi 13914836	/dataset=dbest	/taxon=960...	1298	0.0
gi 2539160	/dataset=dbest	/taxon=9606 ...	1037	0.0
gi 3214685	/dataset=dbest	/taxon=9606 ...	1015	0.0
gi 5933458	/dataset=dbest	/taxon=9606 ...	955	0.0
gi 4988948	/dataset=dbest	/taxon=9606 ...	842	0.0
gi 4900594	/dataset=dbest	/taxon=9606 ...	856	0.0
gi 4534604	/dataset=dbest	/taxon=9606 ...	819	0.0
gi 7455087	/dataset=dbest	/taxon=9606...	789	0.0

**EXPRESSION INFORMATION FOR MODULATORY USE:**

library source:

Expression information from BLAST dbEST hits:

gi 10991237	Neuronal precursor cells-teratocarcinoma
gi 14051054	skin
gi 14076211	skin melanotic melanoma, high MDR (cell line)
gi 11251518	muscle rhabdomyosarcoma
gi 13914836	brain neuroblastoma, cell line
gi 2539160	whole brain
gi 3214685	breast
gi 5933458	stomach
gi 4988948	pancreas - adenocarcinoma
gi 4900594	uterus - serous papillary carcinoma, high grade
gi 4534604	brain - anaplastic oligodendroglioma
gi 7455087	colon - moderately-differentiated adenocarcinoma

Tissue source of cDNA clone:

Fetal whole brain

**FIGURE 1B**

1 MRKMLAAVSR VLSGASQKPA SRVLVASRNF ANDATFEIKK CDLHRLEEGP  
 51 PVTTTLTRED GLKYRMMQT VRRMELKADQ LYKQKIIRGF CHLCDGQFLL  
 101 PLTQEACCVG LEAGINPTDH LITAYRAHGF TFTRGLSVRE ILAELTGRKG  
 151 GCAKAKGGSM HMYAKNFYGG NGIVGAQVPL GAGIALACKY NGKDEVCLTL  
 201 YGDGAANQGQ IFEAYNMAAL WKLPCIFICE NNRYGMGTSV ERAAASTDYY  
 251 KRGDFIPGLR VDGMDILCVR EATRFAAAYC RSGKGPIILME LQTYRYHGHS  
 301 MSDPGVSYRT REEIQEVRSK SDPIMLLKDR MVNSNLASVE ELKEIDVEVR  
 351 KEIEDAAQFA TADPEPPLEE LGYHIYSSDP PFEVRGANQW IKFKSVS (SEQ ID NO:2)

**FEATURES:**

**Functional domains and key regions:**

[1] PDOC00005 PS00005 PKC\_PHOSPHO\_SITE  
 Protein kinase C phosphorylation site

Number of matches: 7

1	16-18	SQK
2	70-72	TVR
3	137-139	SVR
4	146-148	TGR
5	282-284	SGK
6	293-295	TYR
7	307-309	SYR

[2] PDOC00006 PS00006 CK2\_PHOSPHO\_SITE  
 Casein kinase II phosphorylation site

Number of matches: 7

1	57-60	TRED
2	137-140	SVRE
3	238-241	TSVE
4	300-303	SMSD
5	310-313	TREE
6	319-322	SKSD
7	338-341	SVEE

[3] PDOC00008 PS00008 MYRISTYL  
 N-myristoylation site

Number of matches: 7

1	110-115	GLEAGI
2	114-119	GINPTD
3	151-156	GCAKAK
4	172-177	GIVGAQ
5	181-186	GAGIAL
6	183-188	GIALAC
7	235-240	GMGTSV

[4] PDOC00009 PS00009 AMIDATION  
 Amidation site

146-149 TGRK

[5] PDOC00016 PS00016 RGD  
 Cell attachment sequence

252-254 RGD

**Membrane spanning structure and domains:**

Helix	Begin	End	Score	Certainty
1	169	189	1.097	Certain

**FIGURE 2A**

**BLAST Alignment to Top Hit:**

>CRA|18000004925454 /altid=gi|387011 /def=gb|AAA60055.1| (J03503)  
pyruvate dehydrogenase E1-alpha precursor [Homo sapiens]  
/org=Homo sapiens /taxon=9606 /dataset=nraa /length=414  
Length = 414

Score = 846 bits (2163), Expect = 0.0  
Identities = 411/421 (97%), Positives = 411/421 (97%)  
Frame = +3

Query: 18 ETWGHPRRASWVVRSSRCRHCLCFMRKMLAAVSRVLSGASQKPASRVLVASRNFANDATF 197  
Sbjct: 1 ETWGHPRRASWVVRSSRCRHCLCFMRKMLAAVSRVLSGASQKPASRVLVASRNFANDATF 60

Query: 198 EIKKCDLHRLEEGPPVTTVLTREDGLKYRMMQTVRRMELKADQLYKQKIIRGFCHLCDG 377  
Sbjct: 61 EIKKCDLHRLEEGPPVTTVLTREDGLKYRMMQTVRRMELKADQLYKQKIIRGFCHLCDG 120

Query: 378 QFLLPLTQEACCVGLEAGINPTDHLITAYRAHGFTFTRGLSVREILAEITGRKGGCAK 557  
Sbjct: 121 Q-----EACCVGLEAGINPTDHLITAYRAHGFTFTRGLSVREILAEITGRKGGCAK 173

Query: 558 GGSMMHYAKNFYGGNGIVGAQVPLGAGIALACKYNGKDEVCLTYGDGAANQGQIFEAYN 737  
Sbjct: 174 GGSMMHYAKNFYGGNGIVGAQVPLGAGIALACKYNGKDEVCLTYGDGAANQGQIFEAYN 233

Query: 738 MAALWKLPCIFICENNRYGMGTSVERAAASTDYKRGDFIPGLRVDGMDILCVREATRFA 917  
Sbjct: 234 MAALWKLPCIFICENNRYGMGTSVERAAASTDYKRGDFIPGLRVDGMDILCVREATRFA 293

Query: 918 AAYCRSGKGPILMELQTYRYHGHSMSPGVSYRTREEIQEVRKSDPIMLLKDRMVNSNL 1097  
Sbjct: 294 AAYCRSGKGPILMELQTYRYHGHSMSPGVSYRTREEIQEVRKSDPIMLLKDRMVNSNL 353

Query: 1098 ASVEELKEIDVEVRKEIEDAAQFATADPEPPLEELGYHIYSSDPPFEVRGANQWIKFSV 1277  
Sbjct: 354 ASVEELKEIDVEVRKEIEDPAQFAAADPEPPLEELGYHIYSSDPPFEVRGANQWIKFSV 413

Query: 1278 S 1280  
S  
Sbjct: 414 S 414 (SEQ ID NO:4)

>CRA|18000004920128 /altid=gi|4505685 /def=ref|NP\_000275.1| pyruvate  
dehydrogenase (lipoamide) alpha 1; Pyruvate  
dehydrogenase, E1-alpha polypeptide-1 [Homo sapiens]  
/org=Homo sapiens /taxon=9606 /dataset=nraa /length=390  
Length = 390

Score = 793 bits (2025), Expect = 0.0  
Identities = 389/397 (97%), Positives = 389/397 (97%)  
Frame = +3

Query: 90 MRKMLAAVSRVLSGASQKPASRVLVASRNFANDATFEIKKCDLHRLEEGPPVTTVLTRED 269  
Sbjct: 1 MRKMLAAVSRVLSGASQKPASRVLVASRNFANDATFEIKKCDLHRLEEGPPVTTVLTRED 60

Query: 270 GLKYRMMQTVRRMELKADQLYKQKIIRGFCHLCDGQFLLPLTQEACCVGLEAGINPTDH 449  
Sbjct: 61 GLKYRMMQTVRRMELKADQLYKQKIIRGFCHLCDGQ-----EACCVGLEAGINPTDH 113

**FIGURE 2B**

Query: 450 LITAYRAHGFTFTRGLSVREILAEALTGRKGGCAKAKGGSMHMYAKNFYGGNGIVGAQVPL 629  
 Sbjct: 114 LITAYRAHGFTFTRGLSVREILAEALTGRKGGCAK KGGSMHMYAKNFYGGNGIVGAQVPL 173

Query: 630 GAGIALACKYNGKDEVCLTLYGDGAANQGQIFEAYNMAALWKLPCIFICENNRYGMGTSV 809  
 Sbjct: 174 GAGIALACKYNGKDEVCLTLYGDGAANQGQIFEAYNMAALWKLPCIFICENNRYGMGTSV 233

Query: 810 ERAAASTDYYKRGDFIPGLRVDGMDILCVREATRFAAAYCRSGKGPILMELQTYRYHGHS 989  
 Sbjct: 234 ERAAASTDYYKRGDFIPGLRVDGMDILCVREATRFAAAYCRSGKGPILMELQTYRYHGHS 293

Query: 990 MSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFA 1169  
 Sbjct: 294 MSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFA 353

Query: 1170 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFSVS 1280  
 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFSVS  
 Sbjct: 354 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFSVS 390 (SEQ ID NO:5)

>CRA|18000004938217 /altid=gi|6679261 /def=ref|NP\_032836.1| pyruvate  
 dehydrogenase Elalpha subunit [Mus musculus] /org=Mus  
 musculus /taxon=10090 /dataset=nraa /length=390  
 Length = 390

Score = 783 bits (1999), Expect = 0.0  
 Identities = 382/397 (96%), Positives = 387/397 (97%)  
 Frame = +3

Query: 90 MRKMLAAVSRVLSGASQKPASRVLVASRNFANDATFEIKKCDLHRLEEGPPVTTVLTRD 269  
 Sbjct: 1 MRKMLAAVSRVL+G++QKPASRVLVASRNFANDATFEIKKCDLHRLEEGPPVTTVLTRD 60

Query: 270 GLKYYRMMQTVRRMELKADQLYKQKIIRGFCHLCDGQFLLPLTQEACCVGLEAGINPTDH 449  
 Sbjct: 61 GLKYYRMMQTVRRMELKADQLYKQKIIRGFCHLCDGQ EACCVGLEAGINPTDH 113

Query: 450 LITAYRAHGFTFTRGLSVREILAEALTGRKGGCAKAKGGSMHMYAKNFYGGNGIVGAQVPL 629  
 Sbjct: 114 LITAYRAHGFTFTRGL VR ILAEALTGR+GGCAK KGGSMHMYAKNFYGGNGIVGAQVPL 173

Query: 630 GAGIALACKYNGKDEVCLTLYGDGAANQGQIFEAYNMAALWKLPCIFICENNRYGMGTSV 809  
 Sbjct: 174 GAGIALACKYNGKDEVCLTLYGDGAANQGQIFEAYNMAALWKLPCIFICENNRYGMGTSV 233

Query: 810 ERAAASTDYYKRGDFIPGLRVDGMDILCVREATRFAAAYCRSGKGPILMELQTYRYHGHS 989  
 Sbjct: 234 ERAAASTDYYKRGDFIPGLRVDGMDILCVREAT+FAAAYCRSGKGPILMELQTYRYHGHS 293

Query: 990 MSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFA 1169  
 Sbjct: 294 MSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFA 353

Query: 1170 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFSVS 1280  
 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFSVS  
 Sbjct: 354 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFSVS 390 (SEQ ID NO:6)

FIGURE 2C

**Hmmer search results (Pfam):**

Model	Description	Score	E-value	N
PF00676	Dehydrogenase E1 component	598.5	4e-176	1
PF01579	Domain of unknown function	3.0	2.3	1

**Parsed for domains:**

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF01579	1/1	28	46 ..	153	173 .]	3.0	2.3
PF00676	1/1	66	369 ..	1	327 []	598.5	4e-176

**FIGURE 2D**

```

1 AGTTGTTTCCT TCTAACCCAT TGATTTGTTT AATCATGTAT TTAAGTAGGA
51 CCTATATTTT ACTTGTTCCT TGCTATATCT TCAGTGTGTA GTACAGTGTG
101 TGACACAAAA TCGGTGCTCA ATAATAGGTG TTGGATGAAT GAGCAAATGA
151 ATGAATGAAT TCATATTCAT ATGGCCTACA GAGTTCCTCGT ACATGCACAA
201 CCAATATCAC CACCCCGTGG AGATGACTCC CAAATTAATA TTTTGTAGCAA
251 ATGTTCCAGA CTTACAACCT CAACTTCCCG GGGGACATCT TCAGATAGCT
301 GTGCCACTGC CACCACCAGG TCAACATGTC CCAAACCATT CAGACCAGCT
351 TTTTCTCCTG AGCTGGACAT CTGGCCTCCA ACCTTTTCAT TCTCTTTTAC
401 CTTTCATATT CTATCAGCAG CAGCAGCTGC TGAAATCATA CCATGCAAGT
451 TTCTCACGTC CATCTCTGCC TTTTAATGGC GCCCTCTCAC TCCTTTAAGA
501 AGTTTTCTTC CACTGCAACA CGATCTCTCA GTCCAGAGTC TGGCCCAGTG
551 CCCAAATAT TTTCTAGCT ATGCTGAGAG CTGGTCATGC TTTGAACTTC
601 TGCTTTGAAT ACTTTCAGTG ACACTGGGAG AGAATTATCT CATTGGACCA
651 TTGTCAATGT TAGAAAATTC ATTGTTATGC TGAAATGAAA TGATTTTATT
701 CACACACACA CACACACACA CACAAAATAG CTCTTCCTCC TGAACATGA
751 CTGGCCTGAA AATGTGTGAA GACATATCCA ATCCTCTCTG GTTTTACTGT
801 TCATCCAATT TTCTGTTCTC CTCCTGGCAG GAGGATTATA TTTACCTTG
851 TGAAGCTCAG ACATGGTCGG GTAAGTAGCT CTGGTCCGTG AAAATTGAGA
901 GGAAGTGACA TGTGTCACTT CTGGGCAGAA GCTTTGAGAG CCGGTTTAAA
951 TGATCCCTTT TCTCTTCATC CATGAGACAA GCTAAGTCC AGAGAGAGGG
1001 TGCCACGCTG TGAGGGACCT GTGTTACGAG TACGATGGCT CGCGTCACTT
1051 CAAATCTTG AAATCACTGA AATTTGGAGG TCAGTTGTTA CATCATAACC
1101 CAGCCAATTC GTTTAGCCT GTTTTCTTCC TAACTTCTTT AATCGTTCTT
1151 CATAAGTCAC AATCGCAGCC CCTCACGTT CTGACCACTG TCCCTGGAT
1201 TCCACTCAGT TTAATCATT TCCCCCTTAA AATGTGGAGC CCAAATCTGA
1251 ACCCGGAACC CCAGGTGCAA TCCCACTAGG ACACAACACA ATGGGTTTCT
1301 GAGCCCTTTG ATCCTCTGAA TAGAGCCCCT GTTGTGCTTG GTGTTTTGTC
1351 TCTGTGTGTG CTTTTATCAT CGGCTGAGCC ACGCTGTAA CTCGCAGTGA
1401 CGCTGTGAAC CAATAACTAG AGAAAAAGA TTTTCCCAT TGTCCTCTCG
1451 ACATATATTG GGAAACAAAT TTTTGATCC GCGTTCAGT AGACAGGGCA
1501 GAACTGTCCA ACTGCTACGT GATCTTTTAA AGACAAAGTT AGTGGCAGAC
1551 CATTTACAGA AACCAGATGT TCTGTCTTTT GGCTCTGAGC ATGTGCTTAA
1601 TCTTCATCAT CTAGTGTACT GAACGAGATG TACTGAACGA GGGCTGCAGA
1651 GCTCGCAGCA CGGCAGGAGT AGGCGCTCGG TAGGACGGGG CCTGCACAAC
1701 CTCCCCGTA GTACAGCAGAG CGGAATCTAG GAAGGCTCCT TTCCCGCGGC
1751 GCCCTGGAGG CGGGGGCCCC ACCTTCCAC GCAGGCGCTA TCAAGCCCCG
1801 CCTCCTCACC CGCCCGCGGC GTGGCGTCGG AAAGAGCCCT CAGCCCCTCC
1851 CTCTCTGGCG CTGATACCCA ATGGGCAGCC TCAGGCCCTT AGCGGGGGCG
1901 GGGCACCCCC TGGACGCCGT TCTGGTTGGC CCGCGGCCCG GCGCAGCGCA
1951 TGACGTTATT ACGACTCTGT CACGCCGCGG TGCGACTGAG GCGTGGCGTC
2001 TGCTGGGGCA CTTGAAGGAG ACTTGGGGGC ACCCGCGTCG TGCTCCTTG
2051 GTTGTGAGGA GTCGCCGCTG CCGCCACTGC CTGTGCTTCA TGAGGAAGAT
2101 GCTCGCCGCC GTCTCCGCGG TGCTGTCTGG CGCTTCTCAG AAGCCGGTGA
2151 GACCTCCCGG GCGGGCCGGG ATGGGGCGCG AGTGGGGCTG AGGCGGGGCC
2201 GGAGGGCAGG GCGGGCCAGG CCGGGCCACC CAGAGCGGGG TGGAAGGCGC
2251 CAGGGGAGCC GGGGAGCCTT TACTTCGCCT CCGCGCCCTG CATTCGGTTC
2301 CTGGCCTCGG GAGAAGCGGC ACGGACCGGG ATCACGCCAA GGTCCGTGTG
2351 AACTTCCCCC TTCTCGACAC CCACCTCCCC CCCCCGGGCC CAGCTGTGCG
2401 CCAGGCGAAG TCGGTGTGCT CAAGAGGTGC CTGTTGGGTT ACAGGACACG
2451 GAAAGGGTGG CCTCGGCCTC CTTCGAGTCT CCAATTGACC CCACTCATTT
2501 CGGATCTTCT AACTTAATTT CTCTTGACCG AGAGGCTTTG TAATAGCGTA
2551 GAATCTGGAG ACAGGGTGGC TTCGTTCAA CAGCACCTTC ACCATTGACT
2601 AGCCCTGTGA CCTTGAGCAA GTTTTAAAC GTCCCGGGGA CCCGTTTCC
2651 TAAAATGTTT GCTCGAAGTG GAGTTAATCT CTAAATGGAG ATAAGAGTTA
2701 TCTCTGAAAT GTTATCGGTT ATTAAATGT TATCAGTTAA CTCTAAATG
2751 GAGATAATAA GAGTCCCCAC CTCTTGGGGT TGTCTTGAGG ATTCAACGAG
2801 TGACACGTGT GGAACGATT CCAAATAGCA CCTGGCACAT AATCGATAAC
2851 ATGTGTGTTG AATAGTGTTA TTTATTGAGT CTCCAGTTCG GTATACATTT
2901 CTTGAACACC TGTGCTCAGT TCTGAGGCGG GTTCACAGAA GGTCAGCCTC
2951 TTCAGAAACA AACTTCTCTC TCTTCCCTCT CCCTCAACAT CTGAGCTTTT
3001 CTTGGCAGTG AGTTCAGGAG CGCCGAAGCA GAACTCAGAG GACGCTGCCC
3051 TCCCCCTCCC TTACCTACAC ATTCTTAGGG TACAAGTAGC TAAAGCAAAG
3101 AGCAACGATG CTTGAGGGGT GGGGGGTAGA GTTTAGCACT ATTTCATGGC

```

FIGURE 3A

3151 CTCAGCATT AGAGGTGCCT AACACCTGAG CTAGCATTCT GACCCCCCTA  
3201 GGCACAGTGA GGTCTGTGTA ATTGGTGTA CTGCAGGCCT CGGGATTCTG  
3251 GTATTTCCCC CAGGACTTGA TACCGCTCTA CTTAGTACAG GCAAGAGATT  
3301 GTCAAAAAGGT AAAGAGGTAT GCCCTCTAG GAATCCTGTT GCCTAAAATA  
3351 ATGACAAAAC TGCCGGGTGC GGTGCTCAGG CCTGTAATCC CAGCATTTTG  
3401 GGAGGCTGAG GCAGGTGGAT CACCTGAAGG TCAGAAGTTC GAGATCAGCC  
3451 TGGCCAAACAT GGTGAAACCC CGTCTCTACT AAAAATACAA AATTAGCCGG  
3501 TCGTGGTGGC GGGCTCCTGT AATCCCAGCT ACTCGGGAGG CTGAGGCGGG  
3551 AGAATAGCCT GAACCCGGGA GCGGAGTTTG CAGTGAGCGG AGATCGTGCC  
3601 ATTGCACACT GGCTTGGGCG ACAAGAAGCA AGAACTCCGT ATTTTAAAAA  
3651 AAAAAAAAAA AAAAAAAAAA AAAAGCGTTC CCTTAGGGA TATCTGTGGG  
3701 TAGAGGGCTG TACCGGTAGT TACGGGCTCA GAAACATCCT TCCTTTAGGC  
3751 ACCTGATGTA GGTTCCTTC TTCTTCTGCA AGTCAGGTTT ATTGTTTCCT  
3801 GTATCAGTTT GCAGGGTCCC CCCCCCCCCG CCACCTTACA GTAGGAAGAA  
3851 AATTGAGTTC CAGATATGAA GTCACCTTTG AAAAGTGCCCA GGTATCTTTC  
3901 CACTTGGTGG TGTAAGTCTC TCAGATAATT AGAAGTTTTC TGTGTCACTC  
3951 AACTTTGTCAT GGACTAATTT AGGAAACATT CCTGAAGCTT TTAAGGATAG  
4001 AACTAAAAGT TTCACCTTTT TTTTTTTTAA GGGTGGAATA ATAAACTAAC  
4051 GTGTGACTC TTTGTATTTT GTAATTCTTC ATACTTATGG ATGTCTTTTT  
4101 ACTTAACAT AAGTAACAAA ATAGATCAAC GTTTTAGTTT TTTTATATTA  
4151 TACATGTAAA AAGACATTTT GCATATAAGC CTTTCACAAA AATCTTGACA  
4201 GTAAACAATA AGCAGTGGCT CACCCAAATT AGGCAGACTT ACTGCACTAG  
4251 ACTCCTACCA TCTGTGTGAT ACTCCATGAA GGGAGGGAGA AGGGGAGGGA  
4301 GAAGGGTAGG CAGCTGGTCT GATGGCTGTG ACACAAGATA ATCCCTTAA  
4351 CCTCCCAAGA CGCTGTGTGT TTTTTCCTTT TTTATTCTCC CTGGTTTACT  
4401 TTCGTTTTGT TTGAGACAGG GTCTCTGTGT CACCCAGGCT GGAGTGCAGT  
4451 AGCAGGACAG CTCACTGCAG CCTTAGCCTG CTGGGCTCAA GCGATCCTCC  
4501 TGCCTTAGCC TCCTGAGTAG CTGGGAACAC AGGCATGTGC CACCACCACA  
4551 CCCAGCCAAT TAAAAAATT TTTTTTTTAC TAGAGACATG GTCTTGCTAC  
4601 GTTGCCCACT CTGGTCTCCA TCTCCAGGCT CAAGCAGTCC TCCCACCTCG  
4651 GCCTCCCAA GTGCTGGGAT TACTCTCACT CTCTTAAAC CAGGCAGGTA  
4701 GGGAGATTTA TCTCAGGCTT AAAGATTGCC ATTGTCTCAT CAAAGAGTGT  
4751 TTGGTGTGAA ACTTTGAAAT GAATATCAAG ATTGTGTTTT TATTTTGA  
4801 TAAGGTTTAT AGTTTTTATA GTTCTTATTT CATGGAAGAA GATTGAATGC  
4851 ATTTAAAATG TTATTTTATT GTTTGCATTT CTGTATGGCT CCTTTTGTGA  
4901 GATCTTTACT AGCAATGTTT TGGCTTTATA AGTGGTAGGT AAGAGTTTTA  
4951 ATTTACACTG TTAGAATCTG GAATTTTTGA AACGTTTTTC CTCCTTCACA  
5001 TGAATGGTTC CTATGTATTT AGGAAGTTAA AGTTTTACTT TTTTAAATT  
5051 AATTTTTTTT TTAGGCTGG AATGCAGTGG CACAGTCATA GCTCACTGTA  
5101 GCCTCAGGTG TGTGCCACCA TACCTGACTA ATTTTTTAAT ATTTATTTTT  
5151 GTAGAGATGA GAGTCTCATG TTGCCCAGGC TGGCTTTGAA CTCCTGGCTT  
5201 CAAGTGGTCC TCCCACCTG GCCTCCCAA GTGCTGGGGA TTATAGGTGT  
5251 GAGCCATCAT GCGCGCCTA GTTTTTATTT TTTAAAATT GAGTGGGTTG  
5301 TTCGTGGTCT CTGTCAGAGA GGAATCCCAT TTAACAGAGA ATCTTTTTAT  
5351 GGCTCTCCAG AGAAAATGAA TGGTAACTT ATCTTTTCAA CAAGCTCTCA  
5401 CTCAGAAATG ATACACACAC ACTTCTGATA GGACTTTTAG CTTCTTTAAC  
5451 TTTGTTTCCT TCACTCATAT CAGTGGTTCT TATTTTGTAG ATACACAGTA  
5501 ATGAAGCCAT GGGAGAAAGT ATCTAAGTAG CTTTCTGGCA GTCCTAATCT  
5551 TTGCAGGCGC AAGATTACAG GCGCATGCCA CAGCACTGGG CCCCTTCTTG  
5601 CTCTTTATTG TATAGCATTA TCCTGCCTCA TTGTTTCAAC TCTAGGATTG  
5651 AGAAAGAAGT TACCTTTTCT CTGTACTGTG CGCCTGGCTG GTTTGGACTC  
5701 CTGCCTTCCA AAAACTGCAG TTTCTGTAGT TGTATTGGA AATTTATTTT  
5751 ACAATACAAT AAATTTCTGG CCCCACAAA TATTTATTAA CTGCCAAGAA  
5801 TAACACATCT GTTTGATTGC TAAATATAAC CATTGATTTG CTGTTTCACC  
5851 TTCTCTCAGC TTTACTTCTT CCAAATTC TAAATTTTCT TCACTTTTTT  
5901 TGAGATACAT TAGTGGACTG TCTCTGCCTG TAAGTTAACT GAAACACTGA  
5951 TTCCTAGTAT TTCAGTTGTT TTCCTCCAGC ACTGTCATTG TCTGTGTTTG  
6001 TTGGCTTTGT CCAATAATGG TCTATTGAGG GGTGAAGATA TACGTAATTA  
6051 GCTTTCTGCC TATTGGCTTG TACACTCCAG GGTATACTTG GCAGATCAGT  
6101 CTTAACTCTT CTCACCAAGA TCAGTCCAGT GCTGGATTAG GTAAGGTATG  
6151 AACACATCAG ATGTGCTTTT TATGGAGAAA TCATGTTGGT TTACACGTCA  
6201 GTGTGTGAGA ATGTGCGAGA AGGGAGCTAA AATAGTATGA TAATACTACT  
6251 GGATAAATTT TGTGGTCTAA CCTAAACCTT AGCCATTACA TAGAATACTT

FIGURE 3B



6301	TTGCTGTGAG	CAGGTTTGCT	CAGTTGTAAA	ACTGGAAAGG	AATCATTTC
6351	CACCCCCCGC	CTCCAAGCTT	TTTACCTCCA	AACAGTGACA	GCCACCCAAA
6401	CATCAAGAGA	ACAGTGTTC	AGAGAACATT	TCTACTGGGG	CTTCAGGAGG
6451	AGCCTGTCCA	AGATTTAGGC	TGTTCAAATT	ATAAATTATA	AAACAGCTGG
6501	CTCAAGCCCA	TTGTGTTTAA	GTCAGAGAGT	GCTAAGTATC	TTTTCTTTTG
6551	TCTTGCTCTC	CTAAAGTATT	TATCTCATAC	TTCAATCAAT	TTAAATATT
6601	TTTTCTTACA	GATCCAATTT	GATAGAAAG	TCAAGTTTGC	CTAGAGTGGA
6651	GATTAAATCA	TAGTTTTATT	TGAAGTATAA	TTTTGGCTTG	CTCAAAATGA
6701	ACAGTATCTG	GTTATGACTA	AGAATGGCAT	GAAAAGGCCA	GACGCAGTGG
6751	CTCATGCCTG	CAATCCCAGT	ACTTTGGGAG	GCCAAGGCAG	GTGGATCACC
6801	TGAGGTCAGG	AGTTGGAGAC	CAGCCTGGCC	AACATGGTGA	AACCCCATCT
6851	CTACTAAAAA	TATAAAATTT	AGCCGGGCCG	TGGTGGTGGG	CACCTGTAAT
6901	CCCAGCTACT	CGGGAGACTG	AGACAGGAGA	AATCACTTGA	ACCCGGGAAG
6951	CGGAGGTTGC	AGTGAGCCGA	GATCGCACCA	CTGCACTCCA	GCCTGGGTGA
7001	TAAAAGCAAA	ACTCCGTCTC	AAAACAAACA	AACAAAAGAA	TGGCATAAAC
7051	AGACACAGCT	CACAGATGAT	CTAGTCTCTT	TAGCCACTAA	TTTCATTATA
7101	TTCTCACTAT	AATTTCTTTG	AAAACAAAGG	ATGGGTTTGT	TTTTTGCCCC
7151	TCTTTGCGCT	GCTTGCCCTT	AGATGCGGGA	TAATCCTGTT	TCATTGGCCA
7201	AAGCATGGAT	TCATTTTGGA	GGCCAAGGAA	GATGCAAACA	CAGTGCACAG
7251	GGTGGAAGAG	AAGCCTATGA	ATATGTTGGG	GCTTATTAAA	TTTCCATAAC
7301	TCATTCTGA	TAACGTATTA	TTATACTTTC	CAAAATAGCT	GACAATTAAA
7351	AAGTACTGAT	TTGTTTGTAT	ATTTTGTCT	TTTAAGGCAA	GCAGAGTGCT
7401	GGTAGCATCC	CGTAATTTTG	CAAATGATGC	TACATTTGAA	ATTAAGGTAA
7451	GAGTGTTTTA	CTTTGTTAAT	AATTTTTTCA	CAGGTACACT	CTGATATACA
7501	GTTTTACCTT	TAGAATAGAA	CATCTTGATG	TTCATGATTA	GTCATCATTT
7551	TCTTCTAAAT	GTCCAGGATC	AGAAGTTCAG	AGAAGCTTAT	TCAAAAGTTT
7601	GGAATGTAAT	TCAGTGAAAT	ATTTGAATAA	GAAGAGTCTT	AGTTGTTTCT
7651	TTGAAGGTTT	TTTCAACCTA	TAACCTCAGT	GGCTTCTAGG	GGCTTTCAGT
7701	GAAAATCATC	TTAGAAAGAT	TTCTTCCCTC	CAAGCCCAT	CTCATTGCAC
7751	AGTGAGGTTT	ATGGATTTAA	GGAACAGAGG	CGATATGAAG	CATTACTGAT
7801	GTGCTCCTTT	GCAGTTTTTC	AAGTTCAATA	TTATTTGCAA	TGGAGTTAGA
7851	TCTTAGAGTG	GTCAACAGTG	TTTGCAATGT	AGTATGTGGA	GGATAATAAC
7901	TACCTTATTC	CATTTTCAGAA	ATGTGACCTT	CACCGGCTGG	AAGAAGGCCC
7951	TCCTGTCTACA	ACAGTGCTCA	CCAGGGAGGA	TGGGCTCAAA	TACTACAGGA
8001	TGATGCAGAC	TGTACGCCGA	ATGGAGTTGA	AAGCAGATCA	GCTGTATAAA
8051	CAGAAAATTA	TTCTGTGTTT	CTGTCACTTG	TGTGATGGTC	AGGTGAGTGG
8101	TAGGTTTGTG	GTGGAACGTG	GTTATTTAGG	TACTGAAGTA	TGGCTGTGAC
8151	TTATTGGGCT	TTACCTTGCC	ATATGTATCA	GAAGAGTTTG	AGGCTGGTAA
8201	TGTAATTTTC	TTTTATTAT	TTATTTTTTT	GAGACAGTCT	CTCTCTGTCG
8251	CCCAGGTTAG	AGTACAGTGG	TGATCTTGCC	TCAGTGCAGC	CTCTGGTTAG
8301	AGTACAGTGT	GATCTTGGCT	CACTGCAGCC	TCTGTCCACT	GGGCTCAAGC
8351	AATCCTCCCA	CCTCAGCCTC	CCGAGTATGT	GGGACCACAG	GTGCACACCA
8401	ACACACCCAG	CTAATTTTTG	TATTTTTTGG	AGATACGGGG	TTTCACTATG
8451	TTGCCCAGGC	TAGTCTCAAA	CTTCTGGGCT	CAAGTGGTCC	GCCCACCTTG
8501	GCCTCCCAAG	GTGCTAGGAT	TACAGGCGTG	AGCCACTGTG	CCTGGCTGAA
8551	GCCAGTATTT	TAGAATTAAA	AAGTAGAATG	CCAAAACCTG	CTATGAAGCT
8601	TAGGCTAAAG	AATTCATTCA	CACATAACAT	TGCCAGTTT	CTGTACCTGT
8651	TCTTAGAGTT	TTACTATTTT	AAAACCTTCT	GGCACTATGA	TCGCCTGTAC
8701	TGTATATAAT	TTGGAGAGAA	AGGATTAGTT	TGTTTTTTGT	TTTGTGGGCT
8751	TAGGTCAAGG	GTTAGAGTCA	AATACCTACA	AGGGCCAGCC	AGGTAGAATA
8801	AATGAGTGAA	GAAGGCTAGG	TATACAAAAC	AGAAAATGGT	GACAGGGACT
8851	CATGCTGAAC	TGGCACCAGC	ATGCCCTACC	CAGAGGAATG	CCATGACTTG
8901	GTTCCAGCCA	GTTGGTGCCA	TGTGGAAATC	AGGGGTAATG	TTTCCTGTTT
8951	TCCATGTCTA	AGAGAAGGCG	GAAGTCTGGA	TTTTCATGTG	AAATTCCCAG
9001	TGTTTTTAATG	TTGACATCTG	ATGTAGGCTT	TTATTTTAGG	TCATCATACA
9051	GGAGAAAGGA	AGGAAGTGGC	ACATGTGTGG	GTTGCCAGTT	TATTGCTTCT
9101	GGTTTGGGCC	TTCCACTCTG	TATTTTGGGG	GAAAATAGCT	ACTTCTCTCTG
9151	GTTATTAAATG	ACAGGGTCTA	CTAGCCACAC	TATTTCACTG	TGGTCTAGGA
9201	AACGTTTTTA	TTTAGAAACA	TGTATCATAT	TGCCTCATAG	TTTCTCCTTC
9251	CTCTAACACA	GGAAGCTTGC	TGTGTGGGCC	TGGAGGCCGG	CATCAACCCC
9301	ACAGACCATC	TCATCACAGC	CTACCGGGCT	CACGGCTTTA	CTTTCACCCG
9351	GGGCCTTTCC	GTCCGAGAAA	TTCTCGCAGA	GCTTACAGGT	TTGCTGTTGA
9401	TTTACAGAAA	GGGGAATGA	GTGGATTAA	TTTTTAAATA	TCTGTGCATT

FIGURE 3C

9451	AAGATGCTAT	TATGAGTTAA	TATTGTGTTAA	AAATTTTAAG	TTTCTTTTTT
9501	TAACCCCTCTC	TCCTTTGGTG	CTCTGGTACT	TCTGTTGTGC	TCTTGAGTTA
9551	ACTGACCATT	TGTGAAGTTC	TCTGGCCCCT	CAGGTAAG	TTTAAACAG
9601	GTTGGTGCTA	TAAAATCACA	GTAGGTTTGG	TTATCATCA	AGCATGCCAG
9651	AAGAAGTCTA	GCAGTCATAG	AAAGTAAGTT	CGGTTGAAGC	ACTCCATGGT
9701	ATGCAATGTA	AATTCTAGAA	ATCTTCTTAA	TATTCCCTTT	TTCTTTGTCC
9751	CCCGTGACTA	TTTGTTTGTT	TTGGTGGTTT	TTTTTTTTTT	TTTTTTTGA
9801	GACTGTGTCT	CACTCCGTTG	TCCAGGTGGT	GTGCAGTGGT	GTGATCAGGG
9851	CTCACTGCAA	CCTCCACCTC	CCGGGTCAA	GTGATTCTCA	TGCCCTCACC
9901	TCCTGAGTAG	CTGGGACTAC	AGGCATGCAC	CACCACACCT	GGCTAATTTT
9951	TGTATTTTTA	GTAGAGATGG	GGTTTCAACA	TGTTGGCCAG	GCTGGTCTCC
10001	AATCTCTGAG	CTCAGGTGAT	CCACCTGCCT	TGGCCTCCCA	AAGTGTGCTG
10051	GGGTTACAGG	CGTGAGCCAC	CGCACCTGGC	CTGTTTTGTT	TTTTTGAGAC
10101	AGAGTCTCGC	TTTGTTGCCC	AGGCTGGAGT	GCAGTGGCCT	GCCTCAGCCT
10151	CCCAAATGC	TAGGATTACA	GGCGTGAGCC	ACTGTGCCCG	GTCTCTCTCC
10201	TCCTCCTTTT	TTTTTTTTTT	TTTTGAGACA	GAGTTTCACT	CTTTCACCCA
10251	GGCTGGAGTG	GCTGGAGTGA	AGTGGTATGA	TTTTGGCTCA	CTGCAGCCTC
10301	CGCCCCCCGG	CTTCAAGCAA	TTCTCTCGCC	TCAGCCTCCT	GAGTAGCTAG
10351	GATTATAGGT	GCCCAACCAC	CACACCTGGC	TAATTTCTGT	ATTTTATAGTA
10401	GAGACCAGGT	TTCACCATGT	TGGCCAGGCT	GGTCTTGAAC	TCTTGACCTC
10451	AGGTGATCCA	CCCTCTTCGG	CCTCCAAAA	TGTTAGGATT	ACAGGCGTGA
10501	GCCGCCGTGC	CCGGCCCTCC	TTGACTCTTG	AACTATGGTT	GTCCCTCTAT
10551	ATATCCAGGG	GATTGGTTCT	AGGACCCTCG	AGTATACAAA	AATCCTCAAA
10601	TACTCAAGTC	CCAAAGTCAG	CCTTCCATAT	CTTCGGGTTT	GCATCCTGAG
10651	AATATTCTAT	TTTCAATACA	TGTGTGGCTG	AAAAAAAATC	TGTGTATAAG
10701	TGTACCTGTG	CAGTTCAAAC	CCTGTTCAAG	GATTGAATAT	ATTTAGTGTA
10751	CTAGTATAGG	AGAGGTCTTA	AGATGTTTGT	AACTGGCCAG	AAAACCCAGA
10801	AAAGTCCAGG	GTATCATCTG	GATGGAACAT	CTGAAGGAAA	CTAAGTGACT
10851	AGAGAGTAGG	AAAAGCTGGA	AAGGTTGAAG	CACATGGAAC	TAGTGAAAGG
10901	ACAAGGAGAA	ACATGTGTTT	GCCTGGAGGG	ACAGGTACTT	AGACGACTGA
10951	ACTGGCCTCT	GTGTTCTAAT	GGTTGAGCCT	CAGAGTACAT	ATTTGGGGTG
11001	CGGTTTGGTT	TGCTTTGTAG	AGTTGGTTTG	TTCTGCACAT	GTGTATGTTT
11051	TGCCATTTCC	AGGACGAAAA	GGAGGTTGTG	CTAAAGGGAA	AGGAGGATCG
11101	ATGCACATGT	ATGCCAAGAA	CTTCTACGGG	GGCAATGGCA	TCGTGGGAGC
11151	GCAGGTAGT	AAGGACGAGG	ATTGTGTGCT	GCTTTAGATT	TGGCCCTGGA
11201	CTTTGTCTTG	AAAAACCTTT	CACAGCCCCA	GACAACCTTT	CCTGAAGCTA
11251	GTACAGCCAT	GTGCTGCACA	GTGACGCTTT	GGTCAATGTC	GCATATATGA
11301	TGTTGGACCC	ATAAGATTAT	AATGGAGCTG	AAAAATTCCT	GTGCGCTAGT
11351	GATGTTGTAG	TGGCACAACA	CATTACCTTT	TCTACGTTTA	GGTACACAAA
11401	TATTTTGCCT	ACAGGATTCA	GTAGAGTCAC	ATGCTGTGCA	GGGTTGTAGC
11451	CTAGGAGCAG	TAGGCTCTAC	TATACAGCCT	AGGTGTGCAG	TGGGCTGTAC
11501	CATCTAGGTT	CGTGCAATAC	AGTATGGTGT	TCACATGACA	AAATCGCCTA
11551	GTGATGCAAT	TCTGAGAATA	TATCCCTGTT	GTAAAGTGAC	GCGTGACTAT
11601	TTTGGGGGGCT	TGGTTTGCTT	TTAAAGACCT	AGTGCTTCAT	ATCTTACCGT
11651	TTGAGAGATG	AGTAGATTGT	GATGGTGATT	TATAATGTTT	CCTTTTAGGT
11701	GTCTGCTGTT	TTATAAGTAA	GCAGGAACCT	CTAGCAGTGG	AGCCATACCT
11751	TCCCCTTCCT	ATTTATATTT	CAGTACATTA	ATTGCTTTAT	CTTGTCAACT
11801	TCATTTTGGG	GTCCTTGTTT	TCATCAGTTA	GTGAATGATG	AAGAATTAAC
11851	AGCACAAAAT	TATATCCGGA	CTGTTTCTTT	TCCTTTCTAA	TATATTAAGA
11901	TTCTATTATG	TGTTGTTTTT	TTTAAACCT	AGGTTTATT	TTTCCTTTTG
11951	AAATGGAGTC	TTGCTCAGCC	GCCCAGGCTG	GAGCAGTGGT	GTAATCTCAG
12001	CTCACTGCAA	CCTCCACCCC	CGGGTTCAAG	CAATTCTCCT	GCCTCAGCCT
12051	CCCGAGTAGC	TGGGAATATA	GTTACGTGCC	ACCATGCCCC	ACCATTTTTT
12101	GTATTTTTAG	TAGAGACGGG	GTTTCACCAT	CTTGTCCAGG	ATGGTCTCGA
12151	TCTGTGGACC	TCGTGATCTG	CCCAAAGTGC	TGGGATTACA	GGCGTGAGCC
12201	ACCACGCCCG	GCCAGGTTTT	ATTTTTTAAC	TCTTGAATGC	AGAAATGTTA
12251	GTGCTTACTG	GTTAAATAG	AACATAGTAT	TTATATATTA	CTTTAGTGCT
12301	TTATTGAAAA	TATCGAGGT	GGGATAAACA	GAGAGATAGG	GTTGGAAGGA
12351	GAGTTTGTAG	CAGCAGTGTA	ATTCTGTGT	CAGATTCTGG	CCAGGAGTGA
12401	AAATGCAGGG	CATTAAATTAG	TATCTCCCT	CATGGATTTC	TGTGGTTTCT
12451	TTCTCGGTTG	TCCTTAATGT	TAGGTGCCCC	TGGGCGCTGG	GATTGCTCTA
12501	GCCTGTAAGT	ATAATGGAAG	AGATGAGGTC	TGCCTGACTT	TATATGGCGA
12551	TGGTGCTGCT	AACCAGGTAA	TTATGTCTCT	TAACCTCCCA	AAAACAGTCT

FIGURE 3D

12601	TATTTTCAAA	GTCTTTAATA	TTTACAGTTG	AATTTCTAAA	GAAGTAGCAT
12651	ATTGCTTATT	AGGTGAAATA	GCAAGTCCTA	TGGCTAGCTC	AAATTTGGTT
12701	GACTTATGGC	CAGATTAGAG	ATTGACCTCT	TAGCGTTGTT	TCACAAGAGA
12751	CTTACGGGGG	CACATTCCTG	TGAAGGAGCT	CACCTTTGCT	CTACATCAGT
12801	GCTTGGCAAA	GGCCCTGTGG	TAAAGGACCT	CCCCACAACC	TATTGCAAAA
12851	CAATACAGAC	CCATTCTCTT	GGATGTCCGG	GCTGGCAGTG	TCAAATTCGG
12901	ATAATAGCGT	CTGAGTCCTA	ACTCAGTTTC	TATGCTTCTC	TTGTTACCGA
12951	GTAATCCCCA	GTCTGTGGCC	AGCACTCTGT	GAAGCCCTGT	TCTAGAGGCT
13001	GATTCTTAGG	TGCTGGTTCA	CTCTGGCTAT	CCAGTGGGCC	TGATAGATTT
13051	CATATTGATC	TTTTTTCCAG	TGTGTTCCCT	ACTGCTAGCA	TGGCCCCAAA
13101	GAAACAAGTA	GTAGTTGGTT	TGTCACCTTC	CTTAGTTGCA	AGAGTATGAT
13151	GCCTGCTACT	TCTCCTCCAC	CACCCACCCC	GCTTTCCTC	ACCACCCAAA
13201	GCTCGGTTTT	AGAAGAGGAG	GCTTTCCTGT	CTTTATGAAA	GCTTTCCTGT
13251	CCAGGCAGAG	CAGCAGCTGT	TAGAGATGAT	GAAGCCTGGA	GAAAGAAGCC
13301	AAATGAAACC	CCTTTTCGTA	ACTACTTCCA	GGGCCAGATA	TTCGAAGCTT
13351	ACAACATGGC	AGCTTTGTGG	AAATTACCTT	GTATTTTCAT	CTGTGAGAAT
13401	AATCGCTATG	AATGGGAAC	GTCTGTTGAG	AGAGCGGCAG	CCAGCACTGA
13451	TTACTACAAG	AGAGGCGATT	TCATTCCCTG	GCTGAGAGTA	AGGACACCTG
13501	TGGTGGGGCC	GGGGCCAAGG	CCAAGGCCAA	GGGTATGTAC	CTTGTGCAGA
13551	CCCTTGACGA	TCTTAGAAAC	ATTGGAGAGT	TTCATTCTCA	TACAGGAGCA
13601	GGTCATGTGA	AAGTAAAATG	GTTTGGGGCA	GTTGGATTCA	TGCTTCGCCC
13651	CTCCCCGTGT	TATTACCAGG	TGGATGGAAT	GGATATCCTG	TGCGTCCGAG
13701	AGGCAACAAG	TCTTGCTGCC	GCCTATNGTA	GATCTGNNNN	NNNNNNNNNN
13751	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
13801	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
13851	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
13901	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
13951	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14001	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14051	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14351	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14401	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14451	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14501	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14551	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14601	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14651	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14701	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14751	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14801	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14851	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14901	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
14951	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15001	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15051	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15351	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15401	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15451	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15501	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15551	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15601	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15651	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
15701	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNCCTT	TTAGTGTTAC

FIGURE 3E

15751	TTCAGATGAT	ATAGGCATAA	GATACATTGG	TTTTGCTGGC	TGTGCTTCTT
15801	TAGGGGGACT	TAAGGGAGAA	AGGCAAGGCA	CATGGATTTC	CTGCTTGGCG
15851	CTCTGATGTC	TCAAAGTCTA	ATTATCACCA	CACACACCAT	CTCTGCTGTC
15901	CCCACCCATG	TAGTATACAG	GAGCCCAAAT	GGGTGGGACA	AGTGACACTT
15951	CTTTAGAACC	TTACATCTAA	ATCAAAGCAG	CAAGCAAAAA	CTTGGCCCCCT
16001	GTTGTCCGTA	ATGCCAGGGA	AGCCATGTGA	CTCACCAGTG	TACGGTTTTT
16051	TAGAAAAGAC	AGAAGCAGTT	ATTACAGAAT	GTTAGGCTGC	GTTCTGGTAT
16101	TTTGAAAGTA	TAACAACAAC	TCTGCCACGC	CTATAGTGAC	ATAAGCATTG
16151	GTATGCCCTT	TTGTTTCAGA	AACACACTTC	TGTATTTCAC	CTCATTGGGA
16201	CAATCCAACC	CCATATCATG	TTTCATCAGC	CCGTCCTTGC	TCTACTGGAA
16251	CTGCTCTTAC	TGATCGATTA	CTACTTTTCC	CTCCCCATAG	TTACCGTACA
16301	CGAGAAGAAA	TTCAGGAAGT	AAGAAGTAAG	AGTGACCCTA	TTATGCTTCT
16351	CAAGGACAGG	ATGGTGAACA	GCAATCTTGC	CAGTGTGGAA	GAATAAAGG
16401	TACAGTCACT	TGTTTCATGGT	GGTTTGAAGG	TTGGCTTTAA	AAGTTGCCAC
16451	CCCTGGGTGG	CCACAGAGTT	TGTGTGGGTT	CCTCCAAGCC	CAGAAAAGTA
16501	TGTCCTGGGA	CATAAATAGT	TCCATAGTTC	CAAAGTCCCT	TGGGGTGGGG
16551	GCTTTTCCTT	TGATTTCTCT	TATTTCAAAT	TGTATTACTC	TTCAGATTTC
16601	AGATTTTGGT	GGACTGTGAA	CCACCATCAC	AGTGGCAAAG	CCCCACAGT
16651	AGTATGGTTC	TTTTTTCCTA	AAAGTATACT	GTGGATTTTT	AATTCATAAA
16701	ATAGATACAC	CCTAGAAATC	TGTNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
16751	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
16801	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
16851	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
16901	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
16951	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17001	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17051	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17351	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17401	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17451	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17501	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17551	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17601	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17651	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17701	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17751	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17801	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17851	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17901	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
17951	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18001	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18051	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18101	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18151	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18201	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18251	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18301	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN
18351	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN	NNNNNNNNNN

(SEQ ID NO:3)

**FEATURES:**

Start: 2090  
Exon: 2090-2146  
Intron: 2147-7386  
Exon: 7387-7446  
Intron: 7447-7918  
Exon: 7919-8092  
Intron: 8093-9240  
Exon: 9241-9388

**FIGURE 3F**

Intron: 9389-11062  
 Exon: 11063-11154  
 Intron: 11155-12473  
 Exon: 12474-12566  
 Intron: 12567-13331  
 Exon: 13332-13487  
 Intron: 13488-13669  
 Exon: 13670-13727  
 Intron: 13728-15920  
 Exon: 15921-16007  
 Intron: 16008-16290  
 Exon: 16291-

**CHROMOSOME MAP POSITION:**  
 Chromosome X

**ALLELIC VARIANTS (SNPs):**

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
1785	G	T	Beyond ORF (5')			
1895	G	A	Beyond ORF (5')			
2118	G	C	Exon	10	R	P
5144	T	C	Intron			
7932	A	G	Exon	44	H	R
8015	C	T	Exon	72	R	C
8063	C	A	Exon	88	R	S
8066	G	A	Exon	89	G	S
9307	C	G	Exon	120	H	D
9349	C	T	Exon	134	R	W
9350	G	A	Exon	134	R	Q
11066	G	A	Exon	148	R	Q
11128	G	A	Exon	169	G	R
11135	A	G	Exon	171	N	S
11143	G	A	Exon	174	V	M
12486	G	C	Exon	182	A	P
12558	G	A	Exon	206	A	T
13376	T	C A	Exon	223	F	F L
13378	C	T	Exon	224	P	L
16233	G	C	Intron			
16354	G	A	Exon	330	R	K
16377	T	G	Exon	338	C	G

Context:

DNA

Position

1785 TCAAGTAGACAGGGCAGAACTGTCCAACCTGCTACGTGATCTTTTAAAGACAAAGTTAGTG  
 GCAGACCATTACAGAAACCAGATGTTCTGTCTTTTGGCTCTGAGCATGCTGCTAATCTT  
 CATCATCTAGTGTACTGAACGAGATGTACTGAACGAGGGCTGCAGAGCTGCAGCACCGGC  
 AGGAGTAGGCGCTCGGTAGGACGGGGCCTGCACAACCTCCCCGGTAGTCAGCAGAGCGGA  
 ATCTAGGAAGGCTCCTTTCCCGCGGCGCCTGGAGGCGGGGCCCCACCTTCCACGCAG  
 [G, T]  
 CGCTATCAAGCCCCGCCTCCTCACCCGCCCGCGGCGTGGCGTCGGAAAGAGCCCTCAGCC  
 CCTCCCTCTCTGGCGCTGATACCCAATGGGCAGCCTCAGGCCTTAGCGGGGGCGGGGCA  
 CCCCCGACGCGGTTCTGGTTGGCCCCGGCGCGCAGCGCATGACGTTATTACGAC  
 TCTGTCACGCCGCGGTGCGACTGAGGCGTGGCGTCTGCTGGGGCACCTGAAGGAGACTTG  
 GGGGCACCCGCGTGGTGCCTCCTGGGTTGTGAGGAGTCGCCGCTGCCGCCACTGCCTGTG

1895 TGCTAATCTTCATCATCTAGTGTACTGAACGAGATGTACTGAACGAGGGCTGCAGAGCTG  
 CAGCACCGGCAGGAGTAGGCGCTCGGTAGGACGGGGCCTGCACAACCTCCCCGGTAGTCA  
 GCAGAGCGGAATCTAGGAAGGCTCCTTTCCCGCGGCGCCTGGAGGCGGGGCCCCACCT

**FIGURE 3G**

TCCCACGCAGGCGCTATCAAGCCCCGCCTCCTCACCCGCCCGCGGCGTGGCGTCGGAAAG  
AGCCCTCAGCCCCTCCCTCTCTGGCGCTGATACCCAATGGGCAGCCTCAGGCCTTTAGCG  
[G, A]  
GGGCGGGGCACCCCCTGGACGCCGTTCTGGTTGGCCCGGGCCCGCGCAGCGCATGACG  
TTATTACGACTCTGTACGCCCGGTCGACTGAGGCGTGGCGTCTGCTGGGGCACCTGA  
AGGAGACTTGGGGGCACCCGCGTCGTGCCCTCCTGGGTTGTGAGGAGTCGCCGCTGCCGCC  
ACTGCCTGTGCTTCATGAGGAAGATGCTCGCCGCCGTCTCCCGCGTGTGTCTGGCGCTT  
CTCAGAAGCCGGTGAGACCTCCCGGGCGGGCCGGGATGGGGCGCGAGTGGGGCTGAGGCG

2118 GGCCTGGCGTCGGAAAGAGCCCTCAGCCCCCTCCCTCTCTGGCGCTGATACCCAATGGGCA  
GCCTCAGGCCTTTAGCGGGGGCGGGGCACCCCCTGGACGCCGTTCTGGTTGGCCCGCGGC  
CCGGCGCAGCGCATGACGTTATTACGACTCTGTACGCCCGGTCGACTGAGGCGTGGC  
GTCTGCTGGGGCACCTGAAGGAGACTTGGGGGCACCCGCGTCGTGCCCTCCTGGGTTGTGA  
GGAGTCGCCGCTGCCGCCACTGCCTGTGCTTCATGAGGAAGATGCTCGCCGCCGTCTCCC  
[G, C]  
CGTGCTGTCTGGCGCTTCTCAGAAGCCGGTGAGACCTCCCGGGCGGGCCGGGATGGGGCG  
CGAGTGGGGCTGAGGCGGGCCGAGGGCAGGGCGGGCCAGGCCGGGCCACCCAGAGCGG  
GGTGAAGGCGCCAGGGGAGCCGGGGAGCCTTTA

5144 TGAATGCATTTAAAATGTTATTTTATTGTTTGCATTTCTGTATGGCTCCTTTTGTGAGAT  
CTTTACTAGCAATGTTTTGGCTTTTATAAGTGGTAGGTAAGAGTTTTAATTTACACTGTTA  
GAATCTGGAATTTTTGAAACGTTTTTCCCTCTTTCACATGAATGGTTCCTATGTATTTAGG  
AAGTTAAAGTTTTACTTTTTTTTAAATTAATTTTTTTTTTAGGCTGGAATGCAGTGGCAC  
AGTCATAGCTCACTGTAGCCTCAGGTGTGTGCCACCATACCTGACTAATTTTTTAAATATT  
[T, C]  
ATTTTTGTAGAGATGAGAGTCTCATGTTGCCCGAGGCTGGCTTTGAACTCCTGGCTTCAAG  
TGGTCCTCCACCCCTGGCCTCCCAAAGTGCTGGGGATTATAGGTGTGAGCCATCATGCC  
GGCCTAGTTTTTATTTTTTAAATTTGAGTGGGTTGTTTCGTGGTCTCTGTGAGAGAGGAA  
TCCCATTTAACAGAGAATCTTTTTATGGCTCTCCAGAGAAAATGAATGGTAAACTTATCT  
TTTCAACAAGCTCTCACTCAGAAATGATACACACACACTTCTGATAGGACTTTTAGCTTC

7932 AAGAGTCTTAGTTGTTTCTTTGAAGGTTCTTTCAACCTATAACTCAGTTGGCTTCTAGGG  
GCTTTTCAGTGAAATCATCTTAGAAAGATTTCTTCCCCAAGCCCCATCTCATTGCACA  
GTGAGGTTTATGGATTTAAGGAACAGAGGCGATATGAAGCATTACTGATGTGCTCCTTTG  
CAGTTTTTCAAGTTCAATATTATTGCAATGGAGTTAGATCTTAGAGTGGTCAACAGTGT  
TTGCAATGTAGTATGTGGAGGATAATAACTACCTTATTCATTTCAGAAATGTGACCTTC  
[A, G]  
CCGGCTGGAAGAAGGCCCTCCTGTCAACAACAGTGCTCACCAGGGAGGATGGGCTCAAATA  
CTACAGGATGATGCAGACTGTACGCCGAATGGAGTTGAAAGCAGATCAGCTGTATAAACA  
GAAAATTATTTCGTGGTTTCTGTCACTTGTGTGATGGTCAGGTGAGTGGTAGGTTTGTGGT  
GGAAGTGTGTTATTTAGGTACTGAAGTATGGCTTGTACTTATTGGGCTTTACCCCTGCCAT  
ATGTATCAGAAGAGTTTGGGCTGGTAATGTAATTTCTTTTATTTATTTATTTTTTTGA

8015 AAAGATTTCTTCCCCAAGCCCCATCTCATTGCACAGTGAGGTTTATGGATTTAAGGAA  
CAGAGGCGATATGAAGCATTACTGATGTGCTCCTTTGCAGTTTTTCAAGTTCAATATTAT  
TTGCAATGGAGTTAGATCTTAGAGTGGTCAACAGTGTGCAATGTAGTATGTGGAGGAT  
AATAACTACCTTATTCATTTTCAAAATGTGACCTTACCCGGCTGGAAGAAGGCCCTCCT  
GTCACAACAGTGCTCACCAGGGAGGATGGGCTCAAATACTACAGGATGATGCAGACTGTA  
[C, T]  
GCCGAATGGAGTTGAAAGCAGATCAGCTGTATAAACAGAAAATTATTTCGTGGTTTCTGTG  
ACTTGTGTGATGGTCAGGTGAGTGGTAGGTTTGTGGTGGAACTGTGTTATTTAGGTACTG  
AAGTATGGCTTGTACTTATTGGGCTTTACCCTGCCATATGTATCAGAAGAGTTTGAGGCT  
GGTAATGTAATTTCTTTTATTTATTTATTTTTTTGAGACAGTCTCTCTGTGCCCCAG  
GTTAGAGTACAGTGGTGATCTTGGCTCACTGCAGCCTCTGGTTAGAGTACAGTGTGATCT

8063 GGATTTAAGGAACAGAGGCGATATGAAGCATTACTGATGTGCTCCTTTGCAGTTTTTCAA  
GTTCAATATTATTGCAATGGAGTTAGATCTTAGAGTGGTCAACAGTGTGCAATGTAG  
TATGTGGAGGATAATAACTACCTTATTCATTTTCAAAATGTGACCTTACCCGGCTGGAA  
GAAGGCCCTCCTGTCAACAACAGTGCTCACCAGGGAGGATGGGCTCAAATACTACAGGATG  
ATGCAGACTGTACGCCGAATGGAGTTGAAAGCAGATCAGCTGTATAAACAGAAAATTATT  
[C, A]  
GTGGTTTCTGTCACTTGTGTGATGGTCAGGTGAGTGGTAGGTTTGTGGTGGAACTGTGTT  
ATTTAGGTACTGAAGTATGGCTTGTACTTATTGGGCTTTACCCTGCCATATGTATCAGAA

FIGURE 3H

GAGTTTGAGGCTGGTAATGTAATTTCTTTTATTTATTTATTTTTTGAGACAGTCTCTC  
TCTGTGCGCCAGGTTAGAGTACAGTGGTGATCTTGGCTCACTGCAGCCTCTGGTTAGAGT  
ACAGTGTGATCTTGGCTCACTGCAGCCTCTGTCCACTGGGCTCAAGCAATCCTCCCACCT

8066 TTTAAGGAACAGAGGCGATATGAAGCATTACTGATGTGCTCCTTTGCAGTTTTCAGTT  
CAATATTATTTGCAATGGAGTTAGATCTTAGAGTGGTCAACAGTGGTTGCAATGTAGTAT  
GTGGAGGATAATAACTACCTTATTCCATTTAGAAATGTGACCTTCACCGGCTGGAAGAA  
GGCCCTCCTGTCAACAGTGTCTACCAGGAGGATGGGCTCAAATACTACAGGATGATG  
CAGACTGTACGCCGAATGGAGTTGAAAGCAGATCAGCTGTATAAACAGAAAATTATTCGT  
[G, A]  
GTTTCTGTCACTTGTGTGATGGTCAGGTGAGTGGTAGGTTGTGGTGGAAGTGTGTTAT  
TAGGTACTGAAGTATGGCTGTACTTATTGGGCTTTACCCTGCCATATGTATCAGAAGAG  
TTTGAGGCTGGTAATGTAATTTCTTTTATTTATTTATTTTTTGAGACAGTCTCTCTCT  
GTGCGCCAGGTTAGAGTACAGTGGTGATCTTGGCTCACTGCAGCCTCTGGTTAGAGTACA  
GTGTGATCTTGGCTCACTGCAGCCTCTGTCCACTGGGCTCAAGCAATCCTCCCACCTCAG

9307 AATGTTGACATCTGATGTAGGCTTTTATTTTAGGTCATCATACAGGAGAAAGGAAGGAAG  
TGGCAGATGTGTGGGTTGCCAGTTTATTGCTTCTGGTTTGGGCCTTCCACTCTGTATTTT  
GGGGGAAAATAGCTACTTTCTCTGGTTATTAATGACAGGGTCTACTAGCCACATATTTT  
ACTGTGGTCTAGGAAACGTTTATTTAGAAACATGTATCATATTGCCTCATAGTTTCTC  
CTTCTCTAACACAGGAAGCTTGCTGTGTGGGCTGGAGGCCGGCATCAACCCACAGAG  
[C, G]  
ATCTCATCACAGCCTACCGGGCTCACGGCTTTACTTTACCCGGGGCCTTTCCGTCCGAG  
AAATTCTCGCAGAGCTTACAGGTTTGTGTTGATTACAGAAAGGGGAAATGAGTGGATT  
AAGTTTTTAAATATCTGTGCATTAAGATGCTATTATGAGTTAATATTTGTTAAAAATTTT  
AAGTTTCTTTTTTAACCTCTCTCCTTTGGTGCTCTGGTACTTCTGTTGTGCTCTTGAG  
TTAACTGACCATTTGTGAAGTTCTCTGGCCCTCAGGTAAAAGTTTAAACAGGTTGGTG

9349 CAGGAGAAAGGAAGGAAGTGGCAGATGTGTGGGTTGCCAGTTTATTGCTTCTGGTTTGGG  
CCTTCCACTCTGTATTTTGGGGGAAAATAGCTACTTTCTCTGGTTATTAATGACAGGGTCT  
TACTAGCCACATATTTCACTGTGGTCTAGGAAACGTTTATTTAGAAACATGTATCAT  
ATTGCCTCATAGTTTCTCCTTCTCTAACACAGGAAGCTTGCTGTGTGGGCTGGAGGCC  
GGCATCAACCCACAGACCATCTCATCACAGCCTACCGGGCTCACGGCTTTACTTTCACC  
[C, T]  
GGGGCCTTTCCGTCCGAGAAATTCTCGCAGAGCTTACAGGTTTGTGTTGATTACAGAA  
AGGGGAAATGAGTGGATTAAGTTTTTAAATATCTGTGCATTAAGATGCTATTATGAGTTA  
ATATTTGTTAAAAATTTAAGTTTCTTTTTTAACCTCTCTCCTTTGGTGCTCTGGTACT  
TTCTGTGTGCTCTTGAGTTAACTGACCATTTGTGAAGTTCTCTGGCCCTCAGGTAAAA  
GTTTAAACAGGTTGGTGCTATAAAATCACAGTAGGTTTGGTTATCATTCAAGCATGCCA

9350 AGGAGAAAGGAAGGAAGTGGCAGATGTGTGGGTTGCCAGTTTATTGCTTCTGGTTTGGGC  
CTTCCACTCTGTATTTTGGGGGAAAATAGCTACTTTCTCTGGTTATTAATGACAGGGTCT  
ACTAGCCACATATTTCACTGTGGTCTAGGAAACGTTTATTTAGAAACATGTATCATA  
TTGCCTCATAGTTTCTCCTTCTCTAACACAGGAAGCTTGCTGTGTGGGCTGGAGGCCG  
GCATCAACCCACAGACCATCTCATCACAGCCTACCGGGCTCACGGCTTTACTTTCACC  
[G, A]  
GGGCCTTTCCGTCCGAGAAATTCTCGCAGAGCTTACAGGTTTGTGTTGATTACAGAAA  
GGGGAAATGAGTGGATTAAGTTTTTAAATATCTGTGCATTAAGATGCTATTATGAGTTAA  
TATTTGTTAAAAATTTAAGTTTCTTTTTTAACCTCTCTCCTTTGGTGCTCTGGTACT  
TCTGTTGTGCTCTTGAGTTAACTGACCATTTGTGAAGTTCTCTGGCCCTCAGGTAAAA  
TTTAAACAGGTTGGTGCTATAAAATCACAGTAGGTTTGGTTATCATTCAAGCATGCCA

11066 TCCTAAGATGTTTGTAACTGGCCAGAAAACCCAGAAAAGTCCAGGGTATCATCTGGATGG  
AACATCTGAAGGAACTAAGTACTAGAGAGTAGGAAAAGCTGGAAAGGTTGAAGCATAT  
GGAACCTAGTGAAAGGACAAGGAGAAACATGTGTTTGCCTGGAGGGACAGGTACTTAGACG  
ACTGAACTGGCCTCTGTGTTCTAATGGTTGAGCCTCAGAGTACATATTTGGGTTGCGGTT  
TGGTTTGTCTTTGTAGAGTTGGTTTGTCTGCACATGTGTATGTTCTGCCATTTCCAGGAC  
[G, A]  
AAAAGGAGGTTGTGCTAAAGGGAAAGGAGGATCGATGCACATGTATGCCAAGAACTTCTA  
CGGGGGCAATGGCATCGTGGGAGCGCAGGTAGTCAAGGACGAGGATTGTGTGCTGCTTTA  
GATTTGGCCCTGGACTTTGTCTTGAAAAACCTTTCACAGCCCCAGACAACCTTTCTGTAA  
GCTAGTACAGCCATGTGCTGCACAGTGACGCTTGGTCAATGTGCATATATGATGTTGG  
ACCCATAAGATTATAATGGAGCTGAAAAATTCCTGTGCGCTAGTGATGTTGTAGTGGCAC

FIGURE 3I

11128 CATCTGAAGGAACTAAGTGACTAGAGAGTAGGAAAAGCTGGAAAGGTTGAAGCACATGG  
AACTAGTGAAAGGACAAGGAGAAACATGTGTTTGCCTGGAGGGACAGGTACTTAGACGAC  
TGAAC TGGCCTCTGTGTTCTAATGGTTGAGCCTCAGAGTACATATTTGGGGTGC GGTTTG  
GTTTGCTTTGTAGAGTTGGTTTGTCTGCACATGTGTATGTTCTGCCATTTCCAGGACGA  
AAAGGAGGTTGTGCTAAAGGGAAAGGAGGATCGATGCACATGTATGCCAAGAACTTCTAC  
[G, A]  
GGGGCAATGGCATCGTGGGAGCGCAGGTAGTCAAGGACGAGGATTGTGTGCTGCTTTAGA  
TTTGGCCCTGGACTTTGTCTTGAAAAACCTTTCACAGCCCCAGACAACCTTTCTGAAGC  
TAGTACAGCCATGTGCTGCACAGTGACGCTTTGGTCAATGTCGCATATATGATGTTGGAC  
CCATAAGATTATAATGGAGCTGAAAAATTCCTGTCGCCTAGTGATGTTGTAGTGGCACAA  
CACATTACCTTTTCTACGTTTAGGTACACAAATATTTGCCTACAGGATTCAGTAGAGTC

11135 AGGAACTAAGTGACTAGAGAGTAGGAAAAGCTGGAAAGGTTGAAGCACATGGAAC TAGT  
GAAAGGACAAGGAGAAACATGTGTTTGCCTGGAGGGACAGGTACTTAGACGACTGAACTG  
GCCTCTGTGTTCTAATGGTTGAGCCTCAGAGTACATATTTGGGGTGC GGTTTG GTTTGCT  
TTGTAGAGTTGGTTTGTCTGCACATGTGTATGTTCTGCCATTTCCAGGACGAAAAGGAG  
GTTGTGCTAAAGGGAAAGGAGGATCGATGCACATGTATGCCAAGAACTTCTACGGGGGCA  
[A, G]  
TGGCATCGTGGGAGCGCAGGTAGTCAAGGACGAGGATTGTGTGCTGCTTTAGATTTGGCC  
CTGGACTTTGTCTTGAAAAACCTTTCACAGCCCCAGACAACCTTTCTGAAGCTAGTACA  
GCCATGTGCTGCACAGTGACGCTTTGGTCAATGTCGCATATATGATGTTGGACCCATAAG  
ATTATAATGGAGCTGAAAAATTCCTGTCGCCTAGTGATGTTGTAGTGGCACAAACACATTA  
CCTTTTCTACGTTTAGGTACACAAATATTTGCCTACAGGATTCAGTAGAGTCACATGCT

11143 AAGTGACTAGAGAGTAGGAAAAGCTGGAAAGGTTGAAGCACATGGAAC TAGTGAAAGGAC  
AAGGAGAAACATGTGTTTGCCTGGAGGGACAGGTACTTAGACGACTGAACTGGCCTCTGT  
GTTCTAATGGTTGAGCCTCAGAGTACATATTTGGGGTGC GGTTTG GTTTGCTTTGTAGAG  
TTGGTTTGTCTGCACATGTGTATGTTCTGCCATTTCCAGGACGAAAAGGAGGTTGTGCT  
AAAGGGAAAGGAGGATCGATGCACATGTATGCCAAGAACTTCTACGGGGGCAATGGCATC  
[G, A]  
TGGGAGCGCAGGTAGTCAAGGACGAGGATTGTGTGCTGCTTTAGATTTGGCCCTGGACTT  
TGCTTTGAAAAACCTTTCACAGCCCCAGACAACCTTTCCTGAAGCTAGTACAGCCATGTG  
CTGCACAGTGACGCTTTGGTCAATGTCGCATATATGATGTTGGACCCATAAGATTATAAT  
GGAGCTGAAAAATTCCTGTCGCCTAGTGATGTTGTAGTGGCACAAACACATTACCTTTTCT  
ACGTTTAGGTACACAAATATTTGCCTACAGGATTCAGTAGAGTCACATGCTGTGCAGGG

12486 TTACAGGCGTGAGCCACCACGCCCCGCCAGGTTTTATTTTTTAACTCTTGAATGCAGAAA  
TGTTAGTGCTTACTGGTTAAAATAGAACATAGTATTTATATATTACTTTAGTGCTTTATT  
GAAAATATCGGAGGTGGGATAAACAGAGAGATAGGGTTGGAAGGAGAGTTTGTAGCAGCA  
GTGTAATTTCTGTGTCAGATTCTGGCCAGGAGTGAATGCAAGGGCATTAAATTAGTATCT  
CCCCCATGGATTCTGTGTTTCTTTCTCGGTTGTCTTAATGTTAGGTGCCCTGGGCTGGGC  
[G, C]  
CTGGGATTGCTCTAGCCTGTAAGTATAATGGAAAAGATGAGGTCTGCCTGACTTTATATG  
GCGATGGTGCTGCTAACAGGTAATTATGTCTCTTAACTTCCCAAAAACAGTCTTATTTT  
CAAAGTCTTTAATATTTACAGTTGAATTTCTAAGAAGTAGCATATTGCTTATTAGGTGA  
AATAGCAAGTCCTATGGCTAGCTCAAATTTGGTTGACTTATGGCCAGATTAGAGATTGAC  
CTCTTAGCGTTGTTTCAAGAGACTTACGGGGGCACATTCTGTGAAGGAGCTCACCTT

12558 CTGGTTAAAATAGAACATAGTATTTATATATTACTTTAGTGCTTTATTGAAAATATCGGA  
GGTGGGATAAACAGAGAGATAGGGTTGGAAGGAGAGTTTGTAGCAGCAGTGTAATTTCTG  
TGTCAGATTCTGGCCAGGAGTGAATATGCAGGGCATTAATTAGTATCTCCCTCATGGAT  
TTCTGTGGTTCCCTTCTCGGTTGTCTTAATGTTAGGTGCCCTGGGCGCTGGGATTGCT  
CTAGCCTGTAAGTATAATGGAAAAGATGAGGTCTGCCTGACTTTATATGGCGATGGTGCT  
[G, A]  
CTAACAGGTAATTATGTCTCTTAACTTCCCAAAAACAGTCTTATTTTCAAAGTCTTTAA  
TATTTACAGTTGAATTTCTAAGAAGTAGCATATTGCTTATTAGGTGAAATAGCAAGTCC  
TATGGCTAGCTCAAATTTGGTTGACTTATGGCCAGATTAGAGATTGACCTCTTAGCGTTG  
TTTCAAGAGACTTACGGGGGCACATTCTGTGAAGGAGCTCACCTTTGCTCTACATCA  
GTGCTTGGCAAAGGCCCTGTGGTAAAGGACCTCCCAACAACCTATTGCAAAACAATACAG

13376 TCCTTACTGCTAGCATGGCCCCAAAGAAACAAGTAGTAGTTGGTTTGTACCTTCCTTAG  
TTGCAAGAGTATGATGCCTGCTACTTCTCCTCCACCACCCACCCCGCTTTCCCTCACCAC

FIGURE 3J



CCAAAGCTCGGTTTTAGAGAGGAGGCTTTCTGTGCTTTATGAAAGCTTTCTGTGCCAGG  
 CAGAGCAGCAGCTGTTAGAGATGATGAAGCCTGGAGAAAGAAGCCAAATGAAACCCCTTT  
 TCGTAACACTCTCCAGGGCCAGATATTCGAAGCTTACAACATGGCAGCTTTGTGGAAATT  
 [T, C, A]  
 CCTTGTATTTTCATCTGTGAGAATAATCGCTATGGAATGGGAACGTCTGTT

13378 CTTACTGCTAGCATGGCCCCAAAGAAACAAGTAGTAGTTGGTTTGTACCTTCCTTAGTT  
 GCAAGAGTATGATGCCTGCTACTTCTCCTCCACCACCCACCCCGCTTTCCCTCACCACCC  
 AAAGCTCGGTTTTAGAGAGGAGGCTTTCTGTGCTTTATGAAAGCTTTCTGTGCCAGGCA  
 GAGCAGCAGCTGTTAGAGATGATGAAGCCTGGAGAAAGAAGCCAAATGAAACCCCTTTTC  
 GTAACACTCTCCAGGGCCAGATATTCGAAGCTTACAACATGGCAGCTTTGTGGAAATTAC  
 [C, T]  
 TTGTATTTTCATCTGTGAGAATAATCGCTATGGAATGGGAA

16233 GTGGGACAAGTGACACTTCTTTAGAACCTTACATCTAAATCAAAGCAGCAAGCAAAAACCT  
 TGGCCCTGTTGTCGGTAATGCCAGGGAAGCCATGTGACTCACCAGTGACGGTTTTCTA  
 GAAAAGACAGAAGCAGTTATTACAGAATGTTAGGCTGCGTTCTGGTATTTTGAAAGTATA  
 ACAACAACTCTGCCACGCCTATAGTGACATAAGCATTGGTATGCCCTTTGTTTCAGAAA  
 CACACTTCTGTATTTACCTCATTGGGACAATCCAACCCCATATCATGTTTCATCACGCC  
 [G, C]  
 TCCTTGCTCTACTGGAAC TGCTCTTACTGATCGATTACTACTTTTCCCTCCCATAGTTA  
 CCGTACACGAGAAGAAATT CAGGAAGTAAGAAGTAAGAGTGACCCTATTATGCTTCTCAA  
 GGACAGGATGGTGAAACAGCAATCTTGCCAGTGTGGAAGAACTAAAGGTACAGTCACTTGT  
 TCAATGGTGGTTTGAAGGTTGGCTTTAAAAGTTGCCACCCCTGGGTGGCCACAGAGTTTGT  
 GTGGGTTCTCCAAGCCCAGAAAGTGATGTCTGGGACATAAATAGTTCCATAGTTCCAA

16354 AAAAGACAGAAGCAGTTATTACAGAATGTTAGGCTGCGTTCTGGTATTTTGAAAGTATAA  
 CAACAACTCTGCCACGCCTATAGTGACATAAGCATTGGTATGCCCTTTGTTTCAGAAAC  
 ACACTTCTGTATTTACCTCATTGGGACAATCCAACCCCATATCATGTTTCATCACGCCG  
 TCCTTGCTCTACTGGAAC TGCTCTTACTGATCGATTACTACTTTTCCCTCCCATAGTTA  
 CCGTACACGAGAAGAAATT CAGGAAGTAAGAAGTAAGAGTGACCCTATTATGCTTCTCAA  
 [G, A]  
 GACAGGATGGTGAAACAGCAATCTTGCCAGTGTGGAAGAACTAAAGGTACAGTCACTTGT  
 CATGGTGGTTTGAAGGTTGGCTTTAAAAGTTGCCACCCCTGGGTGGCCACAGAGTTTGTG  
 TGGGTTCTCCAAGCCCAGAAAGTGATGTCTGGGACATAAATAGTTCCATAGTTCCAAA  
 GTCCCTTGGGGTGGGGGCTTTTCCTTTAGTTTCTCTATTCAAATTTGTATTACTCTTCA  
 GATTTTCAGATTTTGGTGGACTGTGAACCACCATCACAGTGGCAAAGCCCCACAGTAGTA

16377 GAATGTTAGGCTGCGTTCTGGTATTTTGAAAGTATAACAACAACTCTGCCACGCCTATAG  
 TGACATAAGCATTGGTATGCCCTTTGTTTCAGAAACACACTTCTGTATTTACCTCATT  
 GGGACAATCCAACCCCATATCATGTTTCATCACGCGTCCTTGCTCTACTGGAAGTGCTC  
 TTAAGTATGATCGATTACTACTTTTCCCTCCCATAGTTACCGTACACGAGAAGAAATTCAGG  
 AAGTAAGAAGTAAGAGTGACCCTATTATGCTTCTCAAGGACAGGATGGTGAAACAGCAATC  
 [T, G]  
 TGCCAGTGTGGAAGAACTAAAGGTACAGTCACTTGTTCATGGTGGTTTGAAGGTTGGCTT  
 TAAAAGTTGCCACCCCTGGGTGGCCACAGAGTTTGTGTGGGTTCTCCAAGCCCAGAAAG  
 TGATGTCTGGGACATAAATAGTTCCATAGTTCCAAAGTCCCTTGGGGTGGGGGCTTTTC  
 CTTTAGTTTCTCTATTCAAATTTGTATTACTCTTCAGATTTTCAGATTTTGGTGGACTGT  
 GAACCACCATCACAGTGGCAAAGCCCCACAGTAGTATGGTTCTTTTCTCAAAGTAT

FIGURE 3K